# Voice-Identity Processing Deficit
# The Cognitive and Neural Mechanisms of Phonagnosia

**D I S S E R T A T I O N**

zur Erlangung des akademischen Grades

Doctor rerum naturalium

(Dr. rer. nat.)

eingereicht an der

Lebenswissenschaftlichen Fakultät der Humboldt-Universität zu Berlin

Von **M.A., Claudia Roswandowitz**

Präsidentin

der Humboldt-Universität zu Berlin

Prof. Dr.-Ing. Dr. Sabine Kunst

Dekan der Lebenswissenschaftlichen Fakultät

der Humboldt-Universität zu Berlin

Prof. Dr. Bernhard Grimm

Gutachter/innen

1. Prof. Dr. Katharina von Kriegstein
2. Prof. Dr. Sascha Frühholz
3. Prof. Dr. Werner Sommer

Tag der mündlichen Prüfung: 11.08.2017

## Abstract

The voice contains elementary social communication cues: it conveys speech, as well as paralinguistic information pertaining to the emotional state or the identity of the speaker. In contrast to neuroscientific research on vocal-speech and vocal-emotion processing, voice-identity processing has been less explored. This seems surprising, given the day-to-day significance of person recognition by voice. A valuable approach to unravel how voice-identity processing is accomplished is to investigate people who have a selective deficit in voice recognition. Such a deficit has been termed phonagnosia. In the present dissertation, I investigate phonagnosia following brain damage (i.e. acquired phonagnosia), as well as phonagnosia cases without apparent brain lesion (i.e. developmental phonagnosia). I in-depth characterise the underlying cognitive, neural-functional, and neuro-anatomical mechanisms of phonagnosia by means of comprehensive behavioural testing as well as functional and structural magnetic resonance imaging. The findings of this dissertation inform the current model of voice-identity processing by (i) delivering novel evidence of brain regions that crucially contribute to voice-identity processing, and by (ii) emphasising the multistage nature of voice-identity processing. We showed that dysfunction at different cognitive stages results in behaviourally distinct phonagnosia sub-types. Generally, advanced scientific knowledge about voice-identity processing as provided in the current dissertation also propels practical applications such as clinical treatment programs and artificial voice-recognition systems.

## Zusammenfassung

Die Stimme ist reich an grundlegenden Kommunikationselementen: Sie transportiert Sprache, sowie auch paralinguistische Informationen, wie den emotionale Zustand und die Identität des Sprechers. Im Vergleich zur Sprach- und Emotionsverarbeitung ist die Stimmerkennung bei weitem weniger gut erforscht. Dies ist erstaunlich, angesichts der allgegenwärtigen Relevanz der stimmbasierten Personenerkennung. Ein Ansatz neue Erkenntnisse zur Stimmerkennung zu erlangen, ist die Untersuchung von Personen, die ein selektives Defizit in der Stimmerkennung aufweisen. Dieses Defizit wird Phonagnosie bezeichnet. In der vorliegenden Dissertation untersuche ich Personen, bei denen Phonagnosie nach einer Hirnschädigung aufgetreten ist (i.e. erworbene Phonagnosie) und Personen, bei denen Phonagnosie entwicklungsbedingt und nicht durch eine offensichtliche Hirnschädigung bedingt ist (i.e. entwicklungsbedingte Phonagnosie). Die zugrunde liegenden kognitiven, neuro-funktionalen und neuro-anatomischen Mechanismen der Phonagnosie habe ich detailliert mit Hilfe von umfangreichen Verhaltensuntersuchungen sowie funktionaler und struktureller Magnetresonanztomographie charakterisiert. Die Ergebnisse der Dissertation bereichern das derzeitige Stimmerkennungsmodel durch: (i) Neue Erkenntnisse darüber welche Gehirnregionen notwendig sind, um eine Stimme zu erkennen. (ii) Und durch die empirische Bekräftigung der Annahme, dass Stimmerkennung ein mehrstufiger kognitiver Prozess ist, der verschiedene Subformen der Phonagnosie bedingt. Darüber hinaus tragen neue wissenschaftliche Erkenntnisse zur Stimmerkennung maßgeblich dazu bei, praktische Anwendungsbereiche wie klinische Therapieprogramme und künstliche Stimmerkennungprogramme weiterzuentwickeln.

## Eingereichte Einzelarbeiten

**Roswandowitz C**, Kappes C, Obrig H, von Kriegstein K (in revision) Obligatory and facultative brain regions for voice-identity recognition. [Study 1] – manuscript available from first author

**Roswandowitz C**, Mathias SR, Hintz F, Kreitewolf J, Schelinski S, von Kriegstein K (2014) Two Cases of Selective Developmental Voice-Recognition Impairments. Current Biology 24 (19): 2348-2353. [Study 2] – doi: https://doi.org/10.1016/j.cub.2014.08.048

**Roswandowitz C**, Schelinski S, von Kriegstein K (2017) Developmental phonagnosia: Linking neural mechanisms with the behavioural phenotype. NeuroImage 155, S. 97-112 [Study 3] – doi: https://doi.org/10.1016/j.neuroimage.2017.02.064

## Acknowledgements

First, I would like to express my great gratitude to my supervisor Katharina von Kriegstein. Especially, I would like to thank her for her constant and appreciating support and guidance during my scientific development. Without her confidence in my work I would never be where I am now. Also, I would like to thank Hellmuth Obrig for his great supervision, which always increased my level of motivation. Sincere thanks also go to all members of the "Neural Mechanisms of Human Communication" group, IMPRS peers, and colleagues at the Max Planck Institute who accompanied me during my PhD. Foremost, I would like to thank Sonja Schall, Stefanie Schelinski, Claudia Kappes, and Kamila Borowiak not only for their scientific support but also for their friendships that already proved to survive; even though some of them went on for next adventures. I hope it goes on like that! I would like to thank Corrina Maguinness for fruitful discussions on what voice-identity processing really means.

Furthermore, I would like to thank all participants of my studies for their confidence and time they spent. You made my dissertation as unique as it is.

Most of all, I would like to thank my parents and my sister for never questioning the directions I choose in my life so far. For me, this means unconditional love and support. Danke.

## Table of Contents

# 1 Introduction

It is on the humans' daily agenda to socially communicate. This means to signal others either verbally or non-verbally intentions and attitudes. A crucial fundament for successful social communication is the correct recognition of the communication partner's identity; for instance if it is a friend or a stranger. There are many different communication situations, but for example when we talk to a person, besides the face, the voice is usually the most informative source of identity. Dependent on whether we communicate with a friend or a stranger, we adjust the way of communication. For example, when we meet our best friend compared to a new colleague we behave and talk more informal and emotional. It would be even socially inappropriate to communicate with a stranger in the same way we do with a friend or vice versa.

Besides day-to-day situations, voice-identity processing has also a great relevance in the phylogenesis and ontogenesis of the human being (Sidtis and Kreiman, 2012). Across evolution, recognition of familiar voices has always crucially guided human survival as it indicated whether a voice belongs to a friend or a foe even at night and over long distances outside the range of vision (Sidtis and Kreiman, 2012). As for the human phylogenesis, voice-identity processing plays a significant role for human ontogenesis. Right after birth (DeCasper and Fifer, 1980) and even already in utero (Kisilevsky *et al.*, 2003), infants show a preference for their mother's voice as compared to an unfamiliar female voice, which might be a key ingredient for fostering infant/mother bonding.

Given the relevance of voice-identity processing, the existence of brain areas specifically dedicated to the analysis of voice identity (e.g. Belin *et al.*, 2000; von Kriegstein *et al.*, 2003) is not surprising. And that, in turn, makes voice-identity processing so interesting to investigate: It constitutes a unique case helping to elucidate the mysterious puzzle of the brain.

Most of what we currently know about how and where in the brain voice identities are processed comes from investigations on neuro-typical populations. This evidence builds by-and-large the foundation of the current standard model of voice-identity processing.

Although most people master voice-identity processing with ease and remarkable precision, there are people having a deficit in identifying others by their voice. This neuropsychological deficit has been termed phonagnosia (Van Lancker and Canter, 1982). Investigations on people having phonagnosia present a valuable approach to complement findings on neuro-typical participants as they allow more causal interpretations.

To date, investigations on phonagnosia are limited in their quantity. Especially evidence from neuroimaging research in combination with behaviourally well characterised phonagnosia cases is scarce. This might be due to the relatively late emergence of scientific studies investigating phonagnosia as a deficit (Assal *et al.*, 1976; Van Lancker and Canter, 1982) and the very few available phonagnosia cases (Garrido *et al.*, 2009). Thus, the purpose of the current dissertation is to provide novel clinical evidence on voice-identity processing by combining state-of-the-art neuroimaging methods with in-depth behavioural assessment of well-defined samples of phonagnosia participants.

I studied the cognitive and neuronal mechanisms of phonagnosia to test predictions derived from the current voice-identity processing model. I characterised participant groups having either developmental or acquired phonagnosia. Developmental phonagnosia is defined as a selective deficit in voice-identity processing without apparent brain lesions, whereas acquired phonagnosia reflects impaired voice-identity processing after brain damage.

In the present dissertation, the first section provides a systematic overview of empirical research on phonagnosia, and its relation to the current standard model of voice-identity processing. In the following, I derive my research questions from this review of the state-of-the-art literature on phonagnosia. The next section is dedicated to the empirical studies, which I conducted during my PhD. Based on the study outcomes, I propose a revised version of the standard voice-identity processing model. I conclude with an outlook on future research directions indicating the relevance of voice-identity processing also beyond neurocognitive science.

Please note that some parts of the theoretical background and model descriptions will be published in the book chapter Roswandowitz C.*, Maguinness C.*, von Kriegstein K. (in press). Deficits in voice-identity processing: acquired and developmental phonagnosia. In: The Oxford Handbook of Voice Perception (Frühholz, S., Belin, P., ed), * authors contributed equally.

# 2 Standard Model of Voice-Identity Processing

Although voice-identity processing can often appear effortless, recognising voices at the 'individual' level is a challenge for the perceptual and cognitive system. Voice-identity processing can be conceived as a multistage process, which begins with the encoding of the incoming vocal signal and ends with the successful identification of the voice at the level of a specific individual identity. In the following, I will use the term voice-identity processing to refer to the multistage process including (i) the acoustical voice-identity analysis, (ii) the recognition of voice familiarity, and (iii) the retrieved semantic associations evoked by the recognised voice. In Figure 2.1, the standard model of voice perception is presented.

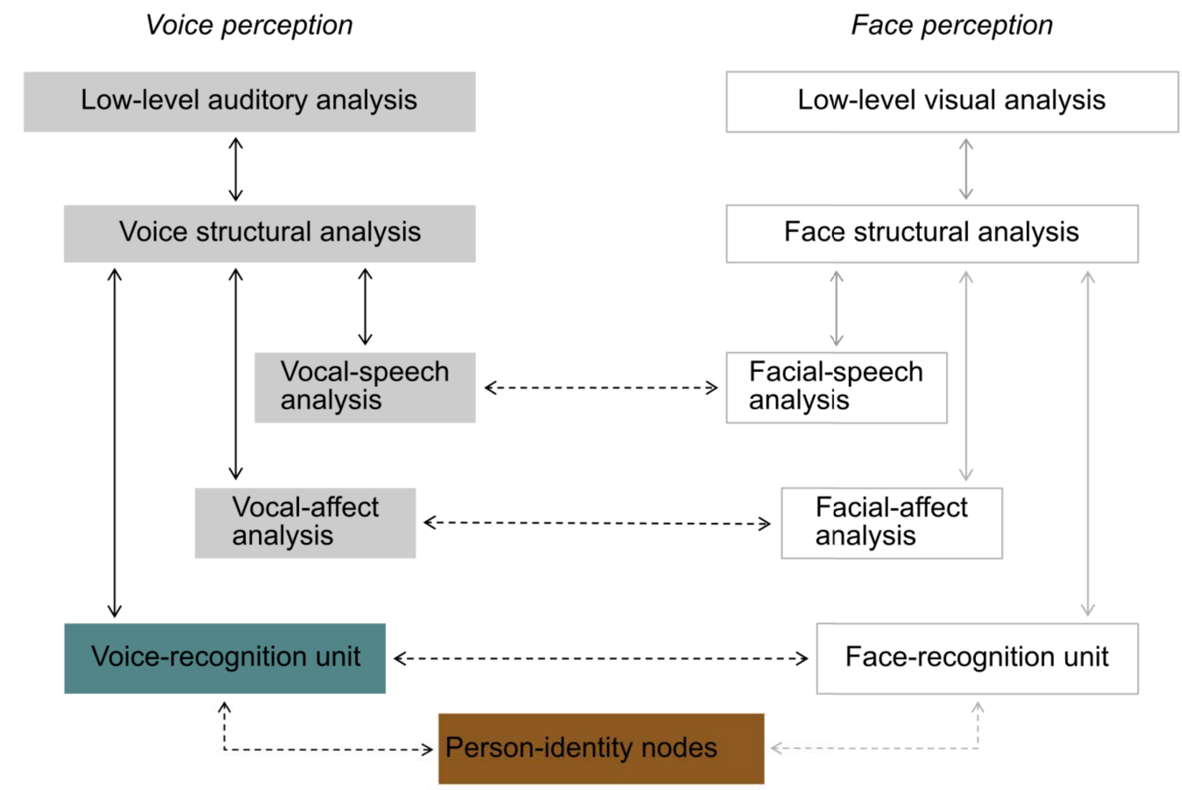


Figure 2.1 Model on voice perception by Belin and colleagues (2004) adapted with permission from Elsevier.

The model has been adapted from Bruce and Young's original model on face perception (right-hand part of the figure) (Bruce and Young, 1986).

In this model, acoustical properties of the auditory signal including vocal and non-vocal sounds are analysed first. This stage is followed by analyses of vocal relevant acoustical information, such as pitch and timbre, at the stage of the 'structural voice analysis'. According to the model, at this stage acoustical properties that are relevant for voice-identity, vocal-speech, and vocal-emotion processing are analysed simultaneously. From this stage onwards, vocal information is further processed in three functionally relatively independent systems responsible (i) for the analysis of vocal-speech information, (ii) for the analysis of vocal-affective information, and (iii) for the analysis of vocal-identity information. If the perceived voice percept closely resembles a stored voice representation a sense of familiarity emerges, i.e. the voice is recognised ('voice recognition unit' VRU; Ellis *et al.*, 1997) (see also prototype encoding of voices, e.g. Lavner *et al.*, 2001). After successful voice recognition, the multimodal person identity node (PIN) is accessed. This can be achieved via both the auditory and visual system. According to the face perception model by Bruce and Young (1986), also Belin and colleagues (2004) assume that the PIN stores individuating information, such as information about the voice, face, and name of the person. The activation of the PIN allows the linkage of further identity-related information. Thus, the listener is able to recall, for instance, the name or the occupation of the speaker. Findings from neuro-typical participants suggest that selective voice-identity processing relies on a core-voice system consisting of temporal voice areas (TVAs), which are located in the bilateral temporal lobes (Belin *et al.*, 2000; von Kriegstein *et al.*, 2003; Pernet *et al.*, 2015). Within the TVAs, acoustical voice-identity processing is assumed to be located in posterior parts of the TVA, i.e. posterior superior temporal gyrus/sulcus (STG/S), middle temporal gyrus/sulcus (MTG/S) as well as Heschl`s gyrus, and planum temporale (von Kriegstein and Giraud, 2004; von Kriegstein *et al.*, 2006; Warren *et al.*, 2006; Formisano *et al.*, 2008; Andics *et al.*, 2010; von Kriegstein *et al.*, 2010; Kreitewolf *et al.*, 2014). Voice-identity recognition, though, is assumed to be supported by the mid to anterior part of TVA (Belin and Zatorre, 2003; von Kriegstein *et al.*, 2003; Andics *et al.*, 2010). Voice-identity processing has been observed bilaterally. Importantly, more carefully controlled fMRI designs, by for instance controlling vocal-speech processing, evidenced BOLD responses more lateralised in the right hemisphere (Belin *et al.*, 2002; von Kriegstein *et al.*, 2003; von Kriegstein and Giraud, 2004). Also clinical studies suggested a right-hemispheric dominance of the TVAs (Assal *et al.*, 1981; Neuner and

Schweinberger, 2000; Lang *et al.*, 2009). In order to associate further individuating information to a familiar voice, supra-modal brain regions belonging to a so called extended system must be accessed. An extended neural system has been first proposed for face-identity processing (Haxby *et al.*, 2000) but is now also implicitly part in models of voice-identity processing (Blank *et al.*, 2014; Perrodin *et al.*, 2015). Candidate areas for this extended system are the temporal pole, precuneus/posterior cingulate, inferior frontal gyrus, amygdala, and regions of other sensory modalities, such as the face-sensitive FFA (fusiform face area) (Shah *et al.*, 2001; von Kriegstein *et al.*, 2005; von Kriegstein and Giraud, 2006; Andics *et al.*, 2010; Latinus *et al.*, 2011; Schall *et al.*, 2013; for a review see Blank *et al.*, 2014).

Although voice- and face-identity processing are assumed to be dissociable mechanisms, the two sensory systems interact with each other, which is indicated by the dashed arrows in Figure 2.1 (Maguinness and von Kriegstein, in press). During voice recognition, neuro-typical participants show structural and functional connectivity between face- and voice-sensitive regions even in the absence of faces (von Kriegstein *et al.*, 2005; Blank *et al.*, 2011; Blank *et al.*, 2014). During person recognition, the recruitment of information across sensory modalities has been ascribed a beneficial rather than a mandatory function (von Kriegstein *et al.*, 2008; Schall *et al.*, 2013).

# 3 Phonagnosia

> „Es war der reine Horror, wenn mich ein Kunde, die Firma, meine Mutter, meine Geschwister, oder ein Freund anriefen!!! Ich brauchte manchmal eine volle Minute, um zu wissen, mit wem ich sprach. Also nach dieser Minute fragte ich den Anrufer, mit wem spreche Ich bitte?“
>
> “It was real horror, when a customer, my mother, my siblings or a friend called!! Sometimes it took me longer than a minute to know who was speaking. After this minute I asked the caller ‘May I ask with who I am speaking to?”
>
> Quote from personal Email conversation (21.12.2016)

The above cited quote vividly pictures that impairments in voice-identity processing affect the quality of every day communication. Phonagnosia originates from the Greek words ‘φώνημα’ or ‘phone’ meaning voice or sound and the term agnosia (αγνώσις). Agnosia is commonly used for conditions, in which the recognition of stimuli is disturbed (Lissauer, 1890; Freud, 1891). Apart from a voice-identity processing deficit, the ability to process other vocal information (e.g., gender, age, and emotion) as well as speech, music, and facial information is largely preserved (Neuner and Schweinberger, 2000; Garrido *et al.*, 2009; Roswandowitz *et al.*, 2014). Moreover, phonagnosia is neither constituted by hearing impairments, nor by a general deficit in cognitive functions. Phonagnosia can occur after brain damage (i.e. acquired phonagnosia) (e.g., Assal *et al.*, 1976; Van Lancker and Canter, 1982; Neuner and Schweinberger, 2000) or in development, in the absence of brain insult (i.e. developmental phonagnosia) (Garrido *et al.*, 2009; Roswandowitz *et al.*, 2014). The disorder has currently two major sub-classifications: apperceptive and associative phonagnosia. In apperceptive phonagnosia, the deficit lies in the perceptual analysis of voice features, while the association of semantic information to a voice remains to be intact (Van Lancker and Kreiman, 1987; Hailstone *et al.*, 2011; Roswandowitz *et al.*, 2014). Apperceptive deficits are commonly tested by means of unfamiliar voice discrimination tests, i.e. whether two sentences were spoken by the same or different speakers. On the other hand, associative phonagnosia is understood as a failure to recognise a voice as familiar (familiarity decision) and to associate semantic information to a voice (semantic association), while the perception of the voice is unaffected (Assal *et al.*, 1981; Van Lancker and Kreiman, 1987; Hailstone *et al.*, 2010; Hailstone *et al.*, 2011; Roswandowitz *et al.*, 2014). To examine associative deficits, familiar voice recognition tests are commonly used. In these tests, listeners first indicate a sense of familiarity towards a voice and then associate semantic knowledge to the familiar voice.

As investigations of malfunction provide causal relationships between brain structure and cognitive functions, phonagnosia may offer a unique instance to study auditory person recognition. Unfortunately, there have been only very few scientific investigations up to now. This might be due to the following factors: (i) phonagnosia became only recently a topic for (neuro)scientific research. The first study on acquired phonagnosia was published in the year 1976 (Assal *et al.*, 1976) and on developmental phonagnosia in the year 2009 (Garrido *et al.*, 2009). (ii) Cases of phonagnosia are rare. This might be due to the fact that the deficit of phonagnosia is still relatively unknown to the public as well as clinicians. Consequently, phonagnosia cases do not recognise their deficit themselves and clinical institutions fail to diagnose phonagnosia; as long as it remains an unknown phenomenon.

## 3.1 Acquired Phonagnosia

The existence of a deficit in identifying voices was first raised by studies on patients suffering from brain damage (e.g., Assal *et al.*, 1976; Van Lancker and Canter, 1982). These studies aimed to characterise the cognitive and neural mechanisms supporting voice-identity processing. Lesion studies on phonagnosia allow a strong interpretation about brain regions obligatory to identify voices.

### 3.1.1 Impaired Voice-Identity Processing after Brain Damage Exists

The first study on acquired phonagnosia was published in 1976 by Assal and colleagues. The authors showed that patients with brain lesions performed significantly worse than neuro-typical controls on discrimination tasks with unfamiliar voices. This provided the first evidence for a voice-identity deficit affecting apperceptive processes after brain injury. In 1987, van Lancker and Kreiman complemented the finding of Assal et al. (1976) by showing that, besides an apperceptive voice-identity processing deficit, also associative processes can be impaired in brain-lesioned patients. They tested patients and neuro-typical controls on unfamiliar voice discrimination and familiar voice recognition. In each of the tasks, patients performed worse compared to controls.

### 3.1.2 Apperceptive and Associative Deficits in Voice-Identity Processing

Having proven evidence for an acquired voice-identity processing deficit, further questions about the nature of the deficit emerged: for example, whether apperceptive and associative voice-identity processes rely on shared or distinct cognitive and neuronal resources. A case report of acquired phonagnosia following brain damage by Assal and colleagues (1981) gave first indication for the dissociation of these mechanisms. The patient RB showed impairments in familiar voice recognition, but his discrimination abilities for unfamiliar voices were mostly preserved. Group studies on brain-damaged patients further confirmed distinct mechanisms at the behavioural and neuronal level (Van Lancker and Kreiman, 1987; Van Lancker *et al.*, 1988; Van Lancker *et al.*, 1989). On the behavioural level, patients showed dissociable performance in both tasks. Patients had impaired unfamiliar voice discrimination and intact familiar voice recognition and vice versa. On the neuronal level, bilateral lesions in the temporal lobe were associated with impaired unfamiliar voice discrimination and lesions restricted to a right-hemispheric parietal system were associated with impaired familiar voice recognition. Also patients with dementia syndromes supported the view of independent apperceptive and associative voice-identity processes. Dementia patients' behavioural pattern was indicative of associative phonagnosia; impaired familiar voice recognition was accompanied by intact unfamiliar voice discrimination (Hailstone *et al.*, 2010; Hailstone *et al.*, 2011).

### 3.1.3 Specificity of Voice-Identity Processing Deficit

The term phonagnosia is broadly used in most of the clinical studies, although it implies a modality-specific deficit that requires many different control tests such as processing of other acoustical information, i.e. vocal speech, vocal emotion, environmental sounds, and music. Control tests should also assess person-recognition abilities by face and name. Most studies were not conclusive as to whether a voice-identity processing disorder may reflect a modality-specific disorder that is unique to vocal identity, i.e. phonagnosia, or a rather multi-modal processing disorder such as a general auditory agnosia or a phonagnosia that is accompanied by prosopagnosia ('face blindness', Bodamer, 1947; McConachie, 1976). A modality-specific disorder would implicate intact face-identity processing and intact other auditory processing abilities.

The relation between voice-identity processing and the identity processing by faces and names as well as other auditory processes was first assessed in a study by Neuner and Schweinberger (2000). They identified 4 patients who suffered from brain damage with selectively impaired familiar voice recognition, while sound, face, and name recognition remained intact. Another study by Lang and colleagues (2009) specifically examined the relation between voice-identity and speech processing. They found that aphasia in patients with left hemispheric lesions was not related to performance in a familiar voice recognition task. In sum, these studies indicate distinct cognitive and neuronal mechanisms for voice-identity processing and face-identity/ speech processing.

### 3.1.4 Brain Lesions Causing Impaired Voice-Identity Processing

Which brain structures are relevant for successful voice-identity processing? In order to answer this question lesion studies present a unique approach. They tell us what damaged brain structure causes a certain behavioural deficit. First studies on acquired phonagnosia have provided evidence on the association between temporal as well as parietal lobe lesions and impaired voice-identity processing (Assal *et al.*, 1976; Assal *et al.*, 1981; Van Lancker and Canter, 1982; Van Lancker and Kreiman, 1987; Van Lancker *et al.*, 1988; Van Lancker *et al.*, 1989; Neuner and Schweinberger, 2000; Lang *et al.*, 2009). However, apperceptive and associative voice-identity processing was linked to differential brain systems. Apperceptive voice-identity processing was associated with lesion in the right and left temporal lobe (Van Lancker *et al.*, 1988; Van Lancker *et al.*, 1989) and associative voice-identity processing only with lesion in the right hemisphere, located in the parietal lobe (Van Lancker *et al.*, 1988; Van Lancker *et al.*, 1989). Recently, more sophisticated lesion-behaviour methods attributed multi-modal person recognition by voice, face, and name to lesions to the anterior temporal lobe (Hailstone *et al.*, 2011).

### 3.1.5 Interim Summary

The summarised clinical studies evidenced the existence of a relatively selective deficit in voice-identity processing after brain damage. On the neuroanatomical level, lesions in the temporal and parietal lobe were associated with impaired voice-identity processing. Supporting the multi-stage nature of voice-identity processing, clinical findings showed dissociable cognitive and neuroanatomical mechanisms for apperceptive and associative voice-identity processing. However, voxel-based evidence for brain regions being selectively involved in voice-identity processing is missing to date.

## 3.2 Developmental Phonagnosia

In the year 2009, the first case has been reported who was indicative of impaired voice-identity processing without apparent brain damage, i.e. developmental phonagnosia (Garrido *et al.*, 2009). What exactly causes developmental phonagnosia is still under debate. According to findings in other agnosias, such as prosopagnosia (Gruter *et al.*, 2007) and amusia (Peretz *et al.*, 2007), a heritable component is suspected. The following section documents the first case of developmental phonagnosia.

### 3.2.1 Case KH

In 2009, Garrido and colleagues reported the first case of developmental phonagnosia, case KH. KH was a 60-year-old female who worked as a successful manager. She presented with a life-long impairment in voice-identity processing. So severe was her deficit that she failed to even recognise the voice of her daughter on the phone. In order to confirm and assess the specificity of her self-report deficit, KH and controls undertook a detailed behavioural battery of vocal-, visual-, and auditory-processing tests. As suspected, compared to controls, KH was significantly impaired in voice-identity processing of famous and newly learned voices (i.e. familiarity judgment and semantic association). Interestingly, KH's ability to discriminate between unfamiliar voices was similar to those of controls. I interpret her behavioural pattern as being suggestive of an associative type of phonagnosia. Garrido and colleagues also showed that KH’s voice-identity processing impairment could not be mediated by a higher order multimodal person-recognition deficit (i.e. face), and, or a general deficit in vocal (i.e. vocal gender, emotion) or auditory processing (i.e. environmental sounds, music). In terms of speech processing, KH’s performance was less clear. She performed well on most tasks, however if the speech signal was embedded in auditory noise, her task performance was below the one of control participants.

### 3.2.2 Interim Summary

The case KH demonstrated the existence of a voice-identity processing deficit that is not a consequence of brain insult. Although KH presented a relatively selective phonagnosia, she was impaired in aspects of speech processing leaving it open whether a selective developmental phonagnosia with intact speech processing exists.

In the meanwhile, three further cases of developmental phonagnosia have been found. After the case KH, I have identified the phonagnosia cases AS and SP. Their behavioural and neural profiles are characterised in Study 2 and 3 of the present dissertation (Roswandowitz *et al.*, 2014; Roswandowitz *et al.*, in press-b). In the following, I introduce the findings of the most recent case, the case AN (Xu *et al.*, 2015).

#### 3.2.3 Case AN

Although AN's report (Xu *et al.*, 2015) was not published at the time when I behaviourally characterised the cases AS and SP (Roswandowitz *et al.*, 2014), for completeness, I briefly summarise AN's behavioural profile here.

Case AN, a 20-year-old female university student stated that she was not particularly aware of her deficit when growing up, as she had not thought that people could recognise an individual without seeing their face. Indeed, AN's face-identity processing was normal (Xu *et al.*, 2015). In terms of voice-identity processing, AN failed to recognise famous voices and to imagine famous voices in comparison to non-voice sounds (Xu *et al.*, 2015). Conversely, AN's ability to process unfamiliar voice identities was intact (i.e. unfamiliar voice matching). Given the dissociation between famous and unfamiliar voice-identity processing, I suggest that her behavioural profile is most likely indicative of an associative type of phonagnosia. Unfortunately, AN's abilities in other auditory tasks such as speech, emotion, and music processing were not formally assessed, leaving open the possibility of a rather more general auditory deficit than a selective impairment in voice-identity processing.

The behavioural characterisation of AN was accompanied by neuro-functional investigations (Xu *et al.*, 2015). The in-depth discussion of AN's neural findings is part of Study 3 where I investigated the neural mechanisms of the cases AS and SP (Roswandowitz *et al.*, in press-b).

# 4 Open Questions Addressed in this Dissertation

As outlined above, the number of studies investigating phonagnosia is limited. The current dissertation aims to investigate further cases of phonagnosia, which, in turn, allows the testing of so far unaddressed model predictions. In the following section, I will introduce the scientific questions of this dissertation and how they relate to the standard model of voice-identity processing.

## 4.1 Do Structures of the Temporal and/or the Parietal Lobe Causally Relate to Voice-Identity Recognition?

Currently, neuroimaging and lesion studies suggest divergent brain regions supporting voice-identity recognition (stage 'Voice recognition unit' in Figure 2.1). Neuroimaging studies, on the one hand, discuss mainly temporal voice areas within the temporal lobe being the most relevant for voice-identity recognition (Belin *et al.*, 2000; von Kriegstein *et al.*, 2003; von Kriegstein and Giraud, 2004; Pernet *et al.*, 2015). On the other hand, the only clinical case study reporting detailed lesion locations showed that lesions in the right inferior parietal lobe cause a deficit in voice-identity recognition (Van Lancker *et al.*, 1988). To date, the standard voice-identity processing model is primarily based on neuroimaging findings. As a consequence, it ascribes a central role of the temporal lobe to the core-voice system while neglecting a potential role of the inferior parietal lobe, as suggested in the clinical study. **Study 1** systematically investigated the causal contribution of lesions in the temporal lobe and the inferior parietal lobe to voice-identity recognition. The evidence for brain structures obligatory for voice-identity recognition tests a core prediction of the voice-identity processing model.

## 4.2 Does a Selective Deficit in Voice-Identity Processing Exist?

The current model on voice perception predicts that voice-identity processing can be dissociated from vocal-speech, vocal-emotion, and face-identity processing. One way to test this prediction is to find people who are exclusively impaired in voice-identity processing while vocal-speech, vocal-emotion, and face-identity processing remain intact. To date, phonagnosia studies were not conclusive to confirm this prediction. The only phonagnosia case who has been tested on all three aspects, i.e. vocal-speech, vocal-emotion, and face recognition (Garrido *et al.*, 2009) evidenced intact face recognition and vocal-emotion but relatively impaired vocal-speech processing. Consequently, within subject evidence for a selective deficit in voice-identity processing is missing to date.

**Study 2** aimed to provide evidence for the existence of selective developmental phonagnosia and to estimate, for the first time, the prevalence of this disorder. By finding cases with impaired voice-identity processing and intact vocal-speech processing, we aimed to test an important model assumption, i.e. dissociation between voice-identity and vocal-speech processing.

## 4.3 What Are the Neural Mechanisms Underlying Developmental Phonagnosia?

Interestingly, the two cases of developmental phonagnosia found in Study 2 showed different behavioural profiles. One case had an apperceptive phonagnosia whereas the other case was characterised by associative phonagnosia.

Previous lesion studies found that apperceptive and associative voice-identity processes are associated with different brain structures. **Study 3** aimed to investigate whether in developmental phonagnosia apperceptive and associative processes also have different neural underpinnings. According to the standard model proposing different contributions of the core-voice and extended system to voice-identity processing, we predicted that impaired acoustical voice-identity processing (i.e. apperceptive phonagnosia) is reflected by a dysfunctional core-voice system and impaired semantic association to a familiar voice (i.e. associative phonagnosia) by a dysconnection between the core-voice and extended system.

# 5 Summary of Empirical Studies

## 5.1 Study 1 - Obligatory and facultative brain regions for voice-identity recognition

To resolve the discrepancy between neuroimaging and lesion findings on brain regions being crucial to recognise human voices, we conducted a voxel-based lesion symptom mapping (VLSM) study. The method of VLSM allows to assess voxel-vise relations between lesioned brain voxels and behavioural test scores (Bates *et al.*, 2003). Study 1 aimed to identify which lesion locations cause a deficit in voice-identity recognition. The patient sample included 58 patients with unselected unilateral brain lesions. 31 patients had lesions in the right hemisphere and 27 patients had lesions in the left hemisphere. For each patient high-resolution structural brain images were available. Lesion types included ischemic stroke ($n$ = 34), traumatic brain injury ($n$ = 7), intracerebral bleeding ($n$ = 6), subarachnoid haemorrhage ($n$ = 6), and tumour excision ($n$ = 4). Patients were tested with a comprehensive behavioural test battery on voice-identity processing. I evaluated patients' abilities to recognise newly learned unfamiliar (voice-name and voice-face test) and familiar voices (famous voice test). The test battery also included tests on face recognition to assess voice-sensitive brain regions. Additionally, patients were tested on abilities to process acoustical voice features, which are important for voice-identity processing, i.e. vocal-pitch and vocal-timbre perception (Lavner *et al.*, 2000; Gaudrain *et al.*, 2009; Sell *et al.*, 2015). All patients additionally took part in a neuropsychological assessment and a pure-tone-audiometry. As I was specifically interested in the functional role of the bilateral temporal lobe and right inferior parietal lobe, a ROI analysis that restricted the VSLM analysis to these brain regions was conducted. The VLSM analysis first investigated whether lesions in the temporal and/or parietal lobe were associated with deficient voice-identity recognition. Additional analyses tested if regions found in the first analysis were selectively sensitive to voice recognition rather than to multi-modal person recognition by voice and face. To test this, the face-recognition score was added as a covariate into the analysis. This allows the analysis of variance in voice recognition (predictor variable) while minimising face recognition deficits as a source of variance (covariate). Effects were considered as significant at $p < 0.01$, cluster-size corrected (1000 permutations). The VLSM analysis revealed three key findings. (i) There was a strong association between lesions in the temporal and right inferior parietal lobe and voice-identity recognition deficits of both newly-learned and familiar voices. Of these two structures, only the right temporal lobe remained significant when controlling for face-recognition performance indicating a high voice sensitivity of the right temporal lobe. (ii) The right inferior parietal lobe was particularly involved in tasks, which required association of voice and face information. (iii) Deficits in recognising newly-learned and familiar voices were associated with distinct brain lesions; newly-learned voice-recognition deficits with lesions in the right temporal and inferior parietal lobe and familiar voices-recognition deficits with left posterior temporal lobe lesions. Importantly, there were no other significant clusters at the whole-brain analysis for any of the tests. With our results we supported a central assumption of the current voice-identity processing model: I provided causal evidence that the right temporal lobe is crucial for successful voice-identity recognition and most likely the key structure of the core-voice system. Our finding on the right inferior parietal lobe also opens new research directions. So far, the inferior parietal lobe is not embedded in current voice-identity processing models. However, I evidenced a substantial, even though facultative, contribution of the inferior parietal lobe to voice-identity recognition. I showed that the right inferior parietal lobe is involved when both voice and face information is available during person recognition.

## 5.2 Study 2 - Two cases of selective developmental voice-recognition impairments

It is a challenging task to find cases with voice-identity processing deficits. By the time when I started my PhD, only one case of developmental phonagnosia, case KH, had been scientifically characterised (Garrido *et al.*, 2009). However, the case KH was not conclusive whether a selective deficit in voice-identity processing exists. Here, I identified and systematically characterised new cases of selective developmental phonagnosia. We developed a novel four-stage screening approach to identify further cases of developmental phonagnosia. The first stage comprised an open access web-based test (http://phonagnosie-test.cbs.mpg.de/) that was designed to self-assess performances in voice recognition. With the phonagnosia online test more than 1,000 data sets from self-selected German individuals were collected. To people who performed 1.5 SD below the laboratory control mean or rated their voice-recognition abilities as poor, a detailed questionnaire was sent. This questionnaire addressed voice-identity processing skills in every day situations as well as a general medical anamneses (*n* = 233). Of the 55 received responses, semi-structured telephone interviews with 10 people who reported voice-recognition problems in the questionnaire were arranged. The fourth stage comprised a detailed behavioural test session in the laboratory that was performed with four cases. Testing included a comprehensive behavioural test battery, an audiometric hearing test, a general neuropsychological assessment. With the innovative screening approach I found two novel cases of selective developmental phonagnosia: AS, a 32-year-old female, and SP, a 32-year-old male; both were otherwise healthy academics, had normal hearing, and showed no pathological abnormalities in brain structure. The two cases performed at least 2 SDs below the level of matched controls (AS's controls *n* = 11, SP's controls *n* = 10) on tests that required learning new voices in association with names, faces, and colours, judging the familiarity of famous voices, and discriminating pitch differences between voices. Intriguingly, only AS failed to discriminate two unfamiliar voices and only SP failed to associate semantic information like the name or face to a familiar recognised voice. This dissociation in task performance indicated different sub-types of phonagnosia. AS's profile suggested an apperceptive phonagnosia where the analysis and integration of acoustic voice features is impaired, whereas the association of semantic information to a voice is intact. SP's profile was indicative of an associative phonagnosia. His deficit in voice-identity processing was characterised by difficulties in associating semantic information to a voice, while the perceptual voice-identity analysis was intact. Besides their distinct voice-identity processing profiles, both cases were remarkably impaired in processing vocal pitch while vocal-timbre processing was intact. I speculate that a pitch deficit per se does not cause phonagnosia but might exacerbates voice-identity processing impairments. In contrast to case KH, speech processing with and without noise was intact in both cases. Also, face recognition, vocal-emotion recognition, and musical abilities were all comparable to controls. The findings confirmed the existence of developmental phonagnosia as a modality- and task-specific impairment. Further, the cases explicitly demonstrated two different behavioural sub-types of phonagnosia, an apperceptive and associative type. On top, our web-based voice recognition test allowed a first rough prevalence estimate. I identified two cases of selective developmental phonagnosia out of over 1000 data sets, suggesting a prevalence of at least 1‰.

## 5.3 Study 3 - Developmental phonagnosia: Linking neural mechanisms with the behavioural phenotype

In Study 3, I examined the neural mechanisms that underlie the behavioural deficit in voice-identity processing found in AS and SP (Study 2) by means of two functional MRI experiments. Studying the neural mechanisms in cases of apperceptive and associative phonagnosia allows to test the prediction that dysfunction in the core-voice system leads to an apperceptive and disconnection between the core-voice and extended system to an associative deficit in voice-identity processing. This in turn would for the first time provide evidence for the assumption that voice-identity processing is accomplished in two neural systems, i.e. a core-voice and extended system. First, I examined AS's and SP's functional brain response profile in the core-voice system using a vocal-sound experiment mostly addressing the perceptual voice analysis. During this fMRI experiment participants passively listened to vocal and non-vocal sounds (Belin *et al.*, 2000). For SP, BOLD responses in the core-voice system for the contrast vocal > non-vocal sounds were comparable to his matched controls ($n$ = 16). This was in accordance with his associative phonagnosia, where perceptual voice-identity processing was intact. In contrast, AS's behavioural profile of poor perceptual voice-identity processing (i.e. apperceptive phonagnosia) was mirrored in reduced response in the core-voice system for the contrast vocal > non-vocal sounds; specifically in the Heschl's gyrus compared to her matched controls ($n$ = 14). Second, functional brain responses in AS and SP in a voice-identity recognition experiment were examined. In this experiment, participants either performed a speaker or a speech recognition task on sentences spoken by different speakers (Schelinski *et al.*, 2016a). What I found was that for the contrast speaker > speech task, AS showed reduced functional brain responses, relative to controls ($n$ = 16), in regions of the core-voice system including the right Heschl's gyrus and planum temporale both extending to the right posterior STS/G. This finding was consistent with her apperceptive deficit. In contrast, in SP, connectivity between the core-voice (i.e. right posterior MTG) and extended system (i.e. amygdala) was altered for the contrast speaker > speech task. As such, his associative deficit in voice-identity processing was likely to arise due to poor connectivity between the core-voice and extended system. In addition, AS and SP showed potential neural compensatory mechanisms for the contrast speaker > speech task matching their distinct behavioural sub-types of phonagnosia: AS recruited more, as compared to her controls, the extended system (i.e. temporal pole, amygdala) as an attempt to compensate her apperceptive deficit, whereas SP recruited more the core-voice system (right mid STG) as a potential compensatory strategy for his associative phonagnosia. Our findings were the first to show that responses in, and connectivity between distinct brain regions, can be associated with discrete behavioural sub-types of phonagnosia. This provided considerable validation of the two-system architecture of voice-identity processing, consisting of a core-voice and an extended system. Our findings demonstrated that cases of phonagnosia, which are apperceptive in nature, may be characterised by atypical functioning within the core-voice system itself (case AS). Additionally, cases of associative phonagnosia may be marked by poor propagation of signals from the (intact) core-voice to the extended system (case SP).

# 6 Study Findings Suggest Revised Version of Voice-Identity Processing Model

In view of the standard voice-identity processing model, the dissertation work supports key assumptions made by the model but also suggests that parts of the model need to be revised (Figure 7.1):

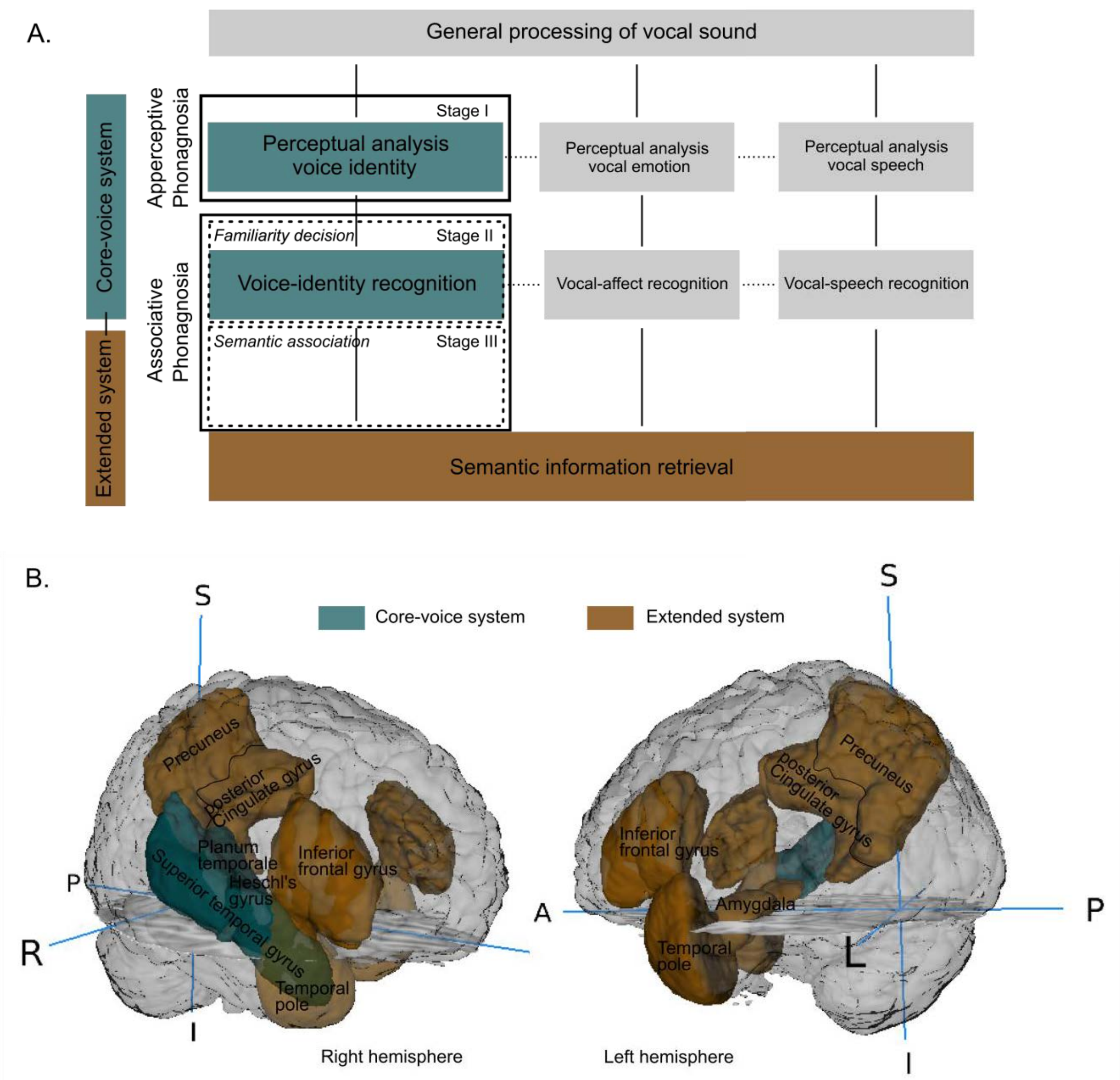


Figure 7.1 Neurocognitive model of voice-identity processing.

A model (adapted from Ellis *et al.* (1997), Belin *et al.* (2004), Blank *et al.* (2014), Perrodin *et al.* (2015)), which describes the cognitive processes (A) and the neural mechanisms (B) supporting voice-identity processing. R= Right, L = Left, S = Superior, I = Inferior, A = Anterior, P = Posterior.

First, by testing brain-damaged patients and cases of developmental phonagnosia I provided evidence which supports a key assumption; namely that the temporal lobe is the central structure of the core-voice system. Second, based on the observation of distinct sub-types of developmental phonagnosia I propose an amendment to the current model by introducing an additional cognitive stage of selective voice-identity processing. In the following, I introduce the revised cognitive and neural model (Figure 7.1A, B), which is based on the standard model (see Figure 2.1) but features the novel empirical evidence of this dissertation.

**Cognitive Mechanisms**

According to the model (Figure 7.1A), the vocal sound undergoes an initial general acoustical processing phase. This processing may be partly shared and partly independent from the processing of other sound sources, including object sounds or music. After this initial phase, voice-identity processing begins.

*Stage I*: Here, the perceptual system analyses complex spectrotemporal acoustical properties of the incoming vocal sound supporting identity processing, i.e. perceptual analysis of voice identity. These properties such as pitch and timbre are merged to create a coherent voice percept, which can be passed on for analysis at later stages of processing. Other features of the vocal sound, which support vocal-affect and speech processing, are also analysed at this stage but are argued to be processed in partly independent but interacting systems (von Kriegstein *et al.*, 2010; Kreitewolf *et al.*, 2014).

Please note, that this stage of processing is novel and thus was not included in the standard model (Figure 2.1). According to the standard model, selective voice-identity processing is only assumed at the stage where the voice is recognised as familiar (see 'Voice recognition unit' in Figure 2.1, and the equivalent stage 'Voice-identity recognition' in Figure 7.1A). Based on the apperceptive phonagnosia case AS (Study 2 and 3), I propose an additional stage of selective voice-identity processing, i.e. a stage for the analysis of acoustical voice-individuating properties (see 'Perceptual analysis voice identity' in Figure 7.1A). The case AS was impaired in perceptual vocal-identity analyses. Importantly, her perceptual deficit resulted exclusively in a deficit in recognising vocal identities but not in recognising vocal speech or vocal emotion. The standard model of voice-identity processing (Figure 2.1), however, does not account for such a behavioural pattern. According to the standard model, deficient perceptual voice analysis would result in a general vocal-sound recognition impairment, including impaired vocal-identity, impaired vocal-speech, and impaired vocal-emotion processing.

*Stage II*: At the stage of voice-identity recognition, a sense of familiarity towards the vocal stimulus is generated. This computation might be based on contrasting the merged vocal percept against a stored prototype (i.e. average) voice representation. The computed acoustical differences between the voice percept and the prototype voice might facilitate the recognition of a unique vocal identity (see prototype encoding of voices; Lavner *et al.*, 2001; Andics *et al.*, 2010; Latinus *et al.*, 2013).

*Stage III*: After the voice has been recognised as familiar the ‘meaning' of the voice is further processed. This can involve the recall of multi-modal semantic information characterising the person’s identity or can involve inferring the intension of the speaker or modulating attention towards an incoming voice (for a review on the extended system in face processing see Haxby *et al.*, 2000).

**Neural Mechanisms**

It has been previously postulated that two neural systems are recruited to successfully recognise who is speaking, i.e. the core-voice and extended system (Blank *et al.*, 2014; Perrodin *et al.*, 2015). With Study 3, we, for the first time, provided evidence for this two-system architecture. We demonstrated a selective contribution of the core-voice system and the connection between the core-voice and extended system to voice-identity processing (Roswandowitz *et al.*, in press-b).

*Core-voice system*: Stages of voice-selective processing (Figure 7.1 A, Stage I, II) are suggested to be supported by a neural system sensitive to voice-identity; i.e. the core-voice system (Figure 7.1A, B). The present dissertation validates the right temporal lobe housing the TVAs as the most likely key structure of this core-voice system. Study 1 showed that lesions in the right temporal lobe caused a selective deficit in voice-identity processing. Specifically, these lesions were located in the posterior STG and planum temporale (Roswandowitz *et al.*, in revision). Study 3 also supported selective voice-identity processing in the temporal lobe: Apperceptive developmental phonagnosia was

characterised by decreased voice-sensitive responses in sub-structures of the temporal lobe, i.e. Heschl's gyrus, planum temporale (Roswandowitz *et al.*, in press-b).
Within the TVA distinct sub-structures have been associated with different cognitive stages. The stage of acoustical voice-identity analysis (Figure 7.1, Stage I) is assigned to posterior and mid regions of the TVA (e.g. von Kriegstein and Giraud, 2004; Warren *et al.*, 2006). The neural response pattern of the apperceptive phonagnosia case, AS (Study 3), supported this assumption. Her perceptual voice-identity processing deficit was related to abnormal processing in brain regions coding acoustical voice-identity analyses, i.e. Heschl's gyrus and planum temporale (Roswandowitz *et al.*, in press-b). Whereas voice-identity recognition (Figure 7.1A, Stage II) is likely supported by anterior/ mid regions of the TVA (Belin and Zatorre, 2003; von Kriegstein *et al.*, 2003; von Kriegstein and Giraud, 2004; Andics *et al.*, 2010; Otto *et al.*, in prep). Please note, for this assumption evidence from phonagnosia cases is missing to date.

*Extended System*: Multi-modal identity information is processed in an extended system (i.e. semantic information retrieval Figure 7.1A, corresponding to the PIN in Figure 2.1), which is proposed to share connections with the core-voice system. Regions concerned with vocal affect and speech recognition may also share connections with this extended system. Among other candidate regions for the extended system (for review see Blank *et al.*, 2014), cases of developmental phonagnosia (Study 3) suggested the anterior temporal lobe and the amygdala as key structures of the extended system (Roswandowitz *et al.*, in press-b).

**Sub-types of Phonagnosia**

The revised model of voice-identity processing which now addresses the multi-stage nature of this process, allows for an explicit introduction of sub-types of phonagnosia (Figure 7.1A). As has been already shown in acquired phonagnosia (Van Lancker and Kreiman, 1987; Van Lancker *et al.*, 1988), findings of Study 2 and 3 evidenced that apperceptive and associative types of phonagnosia can also occur in the developmental variant of this condition (Roswandowitz *et al.*, 2014; Roswandowitz *et al.*, in press-b).

*Apperceptive phonagnosia*: The case AS suggested that apperceptive phonagnosia may emerge due to dysfunction at early stages of processing, i.e. perceptual analysis of voice-identity (Figure 7.1A, Stage I), which are supported by the core-voice system (Figure 7.1A,B). Poor perceptual analysis of the voice may result in a weak representation of voice individuating features, which may impact negatively on later stages of processing (i.e. voice-identity recognition, Figure 7.1A, Stage II).

*Associative phonagnosia*: In contrast, the case of SP suggested that associative phonagnosia may be characterised by malfunction at later stages of voice-identity processing, i.e. poor connectivity between the core-voice system (i.e. voice-identity recognition) and extended system (i.e. semantic information retrieval) (Figure 7.1A, Stage III) (Roswandowitz *et al.*, in press-b). Dysfunction at this stage may underpin cases of *semantic-associative phonagnosia*[1], who are characterised by a deficit in associating semantic information to a voice, which has been successfully perceived and categorised as familiar. Note that direct damage to the extended system, rather than altered connectivity between the core-voice and extended system, would likely result in a multi-modal (i.e. non-voice selective) person-recognition disorder.

A deficit at the stage of voice-identity recognition (Figure 7.1A, Stage II) would likely cause a deficient in familiarity decisions despite a successfully analysed vocal percept. We will call this *familiarity-associative phonagnosia* (Figure 7.1A, Stage II). According to the model, disrupted access to the stored voice-identity representations constrains the ability to judge whether the voice has been encountered before. To date, empirical evidence for familiarity-associative phonagnosia is missing.

[1] The classification of phonagnosia subtypes is informed by the visual agnosia literature (Lissauer, 1890; De Renzi et al., 1991). There, an 'apperceptive' agnosia is consistently categorised as a perceptual processing deficit (Figure 7.1A, Stage I) (Warrington, 1975; De Renzi, 1986; De Renzi et al., 1991). However, much discrepancy exists regarding the definition of 'associative' agnosia in prosopagnosia ('face-blindness'). Classically, associative (prosop)agnosia describes the failure to link an analysed percept to stored multi-modal semantic information (Figure 7.1A, Stage III) (Warrington, 1975; Warrington and Shallice, 1984). However, others have stated that this poor semantic association should be labelled 'amnestic' prosopagnosia and that 'associative' prosopagnosia rather reflects a failure to link the analysed percept to a stored facial representation i.e. impaired familiarity decisions (Figure 7.1A, Stage II) (Fox et al., 2008; Stollhoff et al., 2011; Avidan and Behrmann, 2014). Here, we propose to resolve the discrepancy by using the term 'associative phonagnosia' to refer to a deficit in attributing meaning to the vocal percept. This may arise due to 1) impaired familiarity decisions or 2) impaired semantic association to the vocal identity. To avoid confusion we will call the first familiarity-associative phonagnosia and the second semantic-associative phonagnosia.

# 7 Conclusions and Relevance of Voice-Identity Processing beyond Science

The current dissertation works towards a better understanding of a fundament of human communication; identity recognition of a speaker's voice. The characterisation of the cognitive, neuro-functional, and neuro-anatomical mechanisms underlying phonagnosia notably advances the field of auditory person recognition. This dissertation evidences novel theoretical insights for current models of voice-identity processing. However, replication studies are needed to evidence the generalisability of our findings especially the findings arising from the developmental phonagnosia cases. In the future it would also be desirable to investigate a larger group of developmental phonagnosics instead of single cases. To overcome the difficulty in finding phonagnosia cases standardised test designs may be useful future instruments. However, compared to face-recognition tests, tests for voice-identity processing are often constrained by the language of the listener. Thus testing voice-identity processing beyond a geographical language location with the same vocal stimuli proves difficult. Also, it is important to bear in mind that tests for phonagnosia have to be designed to address the multistage nature of voice-identity processing; tackling perceptual voice-identity analysis, voice-identity recognition, and semantic association (see Figure 7.1A). As a result, standardising the test design as well as the testing procedure may allow for comparison across study findings for phonagnosia. Furthermore, implementing such procedures may also allow for a deeper insight into individual differences in voice-identity processing in the general population. And last but not least, deeper insights into voice-identity processing also hold practical implications for clinical as well as industry-oriented applications. For instance, the dissertation findings constitute a first step towards helping people who suffer from phonagnosia. When I started my work on phonagnosia, the deficit was rarely known to the public. This changed during my PhD. Our scientific work was constantly accompanied by public media attention (i.e. newspaper articles in for instance Spiegel, FAZ, Gehirn und Geist, radio interviews in for instance Deutschlandfunk, DRadioWissen, SWR, MDR_INFO) which resulted in increased public awareness of phonagnosia within German-speaking countries. Based on personal communication with phonagnosia cases, the knowledge that phonagnosia exists has helped these cases tremendously. Some cases reported to be relieved to hear that a deficit in voice-identity processing does not imply a more general impairment in social communication or general cognitive impairments. Training programs to improve voice-identity processing may be important future steps to help phonagnosics. For instance, given the common fundamental vocal-pitch deficit found in cases of developmental apperceptive and associative phonagnosia (Study 2), I speculate that vocal-pitch training might be one promising approach. Improved processing of acoustical properties that make a voice unique, i.e. vocal pitch, might positively affect the stage of perceptual voice-identity processing. In cases of semantic-associative phonagnosia, additional training on semantic recall, such as associating the speaker's name to a voice, may prove beneficial as well. Importantly, such training programs might not only help cases of phonagnosia but also other groups of people who are not able to recognise who is speaking. For instance, people with autisms spectrum disorders suffer, among others symptoms, from voice-identity processing deficits (Schelinski *et al.*, 2016b).

Another cutting edge application of voice-identity processing research is artificial voice-recognition technology. Similar to face-recognition systems, an auditory 'fingerprint' is used for identification purposes in security systems. However, the accuracy of current technologies, especially in real-life situations, is far from satisfactory (e.g. discussion in Kinnunen and Li, 2010). For instance, the human brain recognises a speaker's voice with ease, even when the voice differs in sound, e.g. when the voice is hoarse or nasal due to respiratory illness such as the flu, or when the voice is higher pitched due to emotional arousal. Voice-recognition algorithms struggle with what the human brain masters with ease. To date, artificial voice-recognition systems extract acoustical voice features and compare those to template voices stored in a database by using different models such as Gaussian mixture model, hidden Markov model, or support vector machines (e.g. Reynolds, 2002; Kinnunen and Li, 2010). If an extracted feature matches a template voice, the speaker is successfully recognised. To focus specifically on acoustical voice properties, which distinctively individuate a speaker's voice, could facilitate the recognition accuracy. Until now, psychophysics studies are primarily investigating two individuating voice features: vocal pitch and vocal timbre (Lavner *et al.*, 2000; Gaudrain *et al.*, 2009; Sell et al., 2015, for review see Mathias and von Kriegstein, 2014). These features are considered critical also in artificial voice-recognition systems (Atal, 1976; Rabiner and Juang, 1993; He *et al.*, 1995; Lu and Dang, 2008). However, cognitive studies repeatedly showed that even if voice-recognition performance was decreased after pitch and timbre modification, performance was still above chance level (e.g. Kitamura and Saitou, 2007; Gaudrain *et al.*, 2009; Sell *et al.*, 2015). Accordingly, there might be further and maybe even more relevant properties individuating the human voice. Identifying those properties could propel current voice-recognition algorithms. However, there is currently debate surrounding whether artificial person-recognition systems may also compromise human privacy. Thus, biometrical person-recognition systems, either using the face or the voice as source of identification, should be only applied after careful ethical consideration.

To sum up, the empirical evidence presented in this dissertation supported key model predictions but also suggested a revision to parts of the voice-identity processing model. Further, I highlighted the relevance of voice-identity processing research for people suffering phonagnosia and also for industry-related applications.

## References

Andics A, McQueen JM, Petersson KM, Gal V, Rudas G, Vidnyanszky Z. Neural mechanisms for voice recognition. NeuroImage 2010; 52(4): 1528-40.
Assal G, Buttet J, Thuillard F. [Cerebra hemispheres and auditory perception]. Revue medicale de la Suisse romande 1981; 101(3): 177-85.
Assal G, Zander E, Kremin H, Buttet J. [Voice discrimination in patients with cerebral cortical lesions]. Schweizer Archiv fur Neurologie, Neurochirurgie und Psychiatrie = Archives suisses de neurologie, neurochirurgie et de psychiatrie 1976; 119(2): 307-15.
Atal BS. Automatic recognition of speakers from their voices. Proceedings of the IEEE 1976; 64(4): 460-75.
Bates E, Wilson SM, Saygin AP, Dick F, Sereno MI, Knight RT, *et al.* Voxel-based lesion-symptom mapping. Nature neuroscience 2003; 6(5): 448-50.
Belin P, Fecteau S, Bedard C. Thinking the voice: neural correlates of voice perception. Trends Cogn Sci 2004; 8(3): 129-35.
Belin P, Zatorre RJ. Adaptation to speaker's voice in right anterior temporal lobe. Neuroreport 2003; 14(16): 2105-9.
Belin P, Zatorre RJ, Ahad P. Human temporal-lobe response to vocal sounds. Cognitive Brain Research 2002; 13(1): 17-26.
Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B. Voice-selective areas in human auditory cortex. Nature 2000; 403(6767): 309-12.
Blank H, Anwander A, von Kriegstein K. Direct Structural Connections between Voice- and Face-Recognition Areas. The Journal of neuroscience : the official journal of the Society for Neuroscience 2011; 31(36): 12906-15.
Blank H, Kiebel SJ, von Kriegstein K. How the human brain exchanges information across sensory modalities to recognize other people. Hum Brain Mapp 2014; 36(1): 324-39.
Bodamer J. Die Prosop-Agnosie. Arch Psychiatr Nervenkr Z Gesamte Neurol Psychiatr 1947; 118(1-2): 6-53.
Bruce V, Young A. Understanding face recognition. Br J Psychol 1986; 77(Pt 3): 305-27.
DeCasper AJ, Fifer WP. Of human bonding: newborns prefer their mothers' voices. Science 1980; 208(4448): 1174-6.
Ellis HD, Jones DM, Mosdell N. Intra- and inter-modal repetition priming of familiar faces and voices. BrJPsychol 1997; 88 ( Pt 1)(Journal Article): 143-56.
Formisano E, De Martino F, Bonte M, Goebel R. "Who" Is Saying "What"? Brain-Based Decoding of Human Voice and Speech. Science 2008; 322(5903): 970-3.
Freud S. Zur Auffassung der Aphasien - Eine kritische Studie. Wien: Fischer Taschenbuch; 1891.
Garrido L, Eisner F, McGettigan C, Stewart L, Sauter D, Hanley JR, *et al.* Developmental phonagnosia: a selective deficit of vocal identity recognition. Neuropsychologia 2009; 47(1): 123-31.
Gaudrain E, Li S, Ban VS, Patterson RD. The role of glottal pulse rate and vocal tract length in the perception of speaker identity. Interspeech 2009: 10th Annual Conference of the International Speech Communication Association 2009 2009; 1-5: 152-5.
Gruter M, Gruter T, Bell V, Halligan PW, Horst J, Sperling K, *et al.* Hereditary prosopagnosia: The first case series. Cortex; a journal devoted to the study of the nervous system and behavior 2007; 43(6): 734-49.

Hailstone JC, Crutch SJ, Vestergaard MD, Patterson RD, Warren JD. Progressive associative phonagnosia: a neuropsychological analysis. Neuropsychologia 2010; 48(4): 1104-14.
Hailstone JC, Ridgway GR, Bartlett JW, Goll JC, Buckley AH, Crutch SJ, *et al.* Voice processing in dementia: a neuropsychological and neuroanatomical analysis. Brain 2011; 134(9): 2535-47.
Haxby JV, Hoffman EA, Gobbini MI. The distributed human neural system for face perception. Trends in cognitive sciences 2000; 4(6): 223-33.
He J, Liu L, Palm G. On the use of features from prediction residual signals in speaker identification. Fourth European Conference on Speech Communication and Technology; 1995; 1995.
Kinnunen T, Li H. An overview of text-independent speaker recognition: From features to supervectors. Speech communication 2010; 52(1): 12-40.
Kisilevsky BS, Hains SMJ, Lee K, Xie X, Huang HF, Ye HH, *et al.* Effects of experience on fetal voice recognition. Psychological science 2003; 14(3): 220-4.
Kitamura T, Saitou T. Effects of acoustic modification on perception of speaker characteristics for sustained vowels. Acoustical science and technology 2007; 28(6): 434-7.
Kreitewolf J, Gaudrain E, von Kriegstein K. A neural mechanism for recognizing speech spoken by different speakers. Neuroimage 2014; 91: 375-85.
Lang CJ, Kneidl O, Hielscher-Fastabend M, Heckmann JG. Voice recognition in aphasic and non-aphasic stroke patients. Journal of neurology 2009; 2009/04/09(Journal Article).
Latinus M, Crabbe F, Belin P. Learning-Induced Changes in the Cerebral Processing of Voice Identity. Cerebral cortex 2011.
Latinus M, McAleer P, Bestelmeyer PE, Belin P. Norm-based coding of voice identity in human auditory cortex. Current biology : CB 2013; 23(12): 1075-80.
Lavner Y, Gath I, Rosenhouse J. The effects of acoustic modifications on the identification of familiar voices speaking isolated vowels. Speech Commun 2000; 30(Journal Article): 9-26.
Lavner Y, Rosenhouse J, Gath I. The prototype model in speaker identification by human listeners. IJST 2001; 4(1): 63-74.
Lissauer H. Ein Fall von Seelenblindheit nebst einem Beitrage zur Theorie derselben. Archiv f Psychiatrie 1890; 21(2): 222-70.
Lu X, Dang J. An investigation of dependencies between frequency components and speaker characteristics for text-independent speaker identification. Speech communication 2008; 50(4): 312-22.
Maguinness C, von Kriegstein K. Cross-modal processing of voices and faces in developmental prosopagnosia and developmental phonagnosia. in press.
Mathias SR, von Kriegstein K. How do we recognise who is speaking. Front Biosci (Schol Ed) 2014; 6: 92-109.
McConachie HR. Developmental prosopagnosia. A single case report. Cortex; a journal devoted to the study of the nervous system and behavior 1976; 12(1): 76-82.
Neuner F, Schweinberger SR. Neuropsychological impairments in the recognition of faces, voices, and personal names. Brain Cogn 2000; 44(3): 342-66.
Otto C, Roswandowitz C, Matthias SR, Ragert P, von Kriegstein K. Causal evidence for the right anterior temporal voice region being involved in speaker identification: A tDCS study. in prep.

Peretz I, Cummings S, Dubé M-P. The genetics of congenital amusia (tone deafness): a family-aggregation study. The American Journal of Human Genetics 2007; 81(3): 582-8.
Pernet CR, McAleer P, Latinus M, Gorgolewski KJ, Charest I, Bestelmeyer PE*, et al.* The human voice areas: Spatial organization and inter-individual variability in temporal and extra-temporal cortices. Neuroimage 2015; 119: 164-74.
Perrodin C, Kayser C, Abel TJ, Logothetis NK, Petkov CI. Who is That? Brain Networks and Mechanisms for Identifying Individuals. Trends in cognitive sciences 2015; 19(12): 783-96.
Rabiner LR, Juang B-H. Fundamentals of speech recognition. 1993.
Reynolds DA. An overview of automatic speaker recognition technology. Acoustics, speech, and signal processing (ICASSP), 2002 IEEE international conference on; 2002: IEEE; 2002. p. IV-4072-IV-5.
Roswandowitz C, Kappes C, Obrig H, Von Kriegstein K. Obligatory and facultative brain regions for voice-identity recognition. in revision.
Roswandowitz C, Maguinness C, von Kriegstein K. Deficits in voice-identity processing: acquired and developmental phonagnosia. In: Frühholz S, Belin, P., editor. The Oxford Handbook of Voice Perception; in press-a.
Roswandowitz C, Mathias Samuel R, Hintz F, Kreitewolf J, Schelinski S, von Kriegstein K. Two Cases of Selective Developmental Voice-Recognition Impairments. Current Biology 2014; 24(19): 2348-53.
Roswandowitz C, Schelinski S, von Kriegstein K. Developmental phonagnosia: Linking neural mechanisms with the behavioural phenotype. NeuroImage in press-b.
Schall S, Kiebel SJ, Maess B, von Kriegstein K. Early auditory sensory processing of voices is facilitated by visual mechanisms. Neuroimage 2013; 77: 237-45.
Schelinski S, Borowiak K, von Kriegstein K. Temporal voice areas exist in autism spectrum disorder but are dysfunctional for voice identity recognition. Soc Cogn Affect Neurosci 2016a; 11(11): 1812-22.
Schelinski S, Roswandowitz C, von Kriegstein K. Voice identity processing in autism spectrum disorder. Autism Res 2016b.
Sell G, Suied C, Elhilali M, Shamma S. Perceptual susceptibility to acoustic manipulations in speaker discrimination. Journal of the Acoustical Society of America 2015; 137(2): 911-22.
Shah NJ, Marshall JC, Zafiris O, Schwab A, Zilles K, Markowitsch HJ*, et al.* The neural correlates of person familiarity. A functional magnetic resonance imaging study with clinical implications. Brain 2001; 124(Pt 4): 804-15.
Sidtis D, Kreiman J. In the beginning was the familiar voice: personally familiar voices in the evolutionary and contemporary biology of communication. Integr Psychol Behav Sci 2012; 46(2): 146-59.
Van Lancker D, Kreiman J. Voice discrimination and recognition are separate abilities. Neuropsychologia 1987; 25(5): 829-34.
Van Lancker DR, Canter JG. Impairment of Voice and Face Recognition in Patients with Hemispheric Damage. Brain Cogn 1982; 1(Journal Article): 185-95.
Van Lancker DR, Cummings JL, Kreiman J, B.H D. Phonagnosia; A dissociation between familiar and unfamiliar voices. Cortex; a journal devoted to the study of the nervous system and behavior 1988; 24(Journal Article): 195-209.
Van Lancker DR, Kreiman J, Cummings J. Voice perception deficits: neuroanatomical correlates of phonagnosia. JClinExpNeuropsychol 1989; 11(5): 665-74.

von Kriegstein K, Dogan O, Gruter M, Giraud AL, Kell CA, Gruter T, *et al.* Simulation of talking faces in the human brain improves auditory speech recognition. ProcNatlAcadSciUSA 2008; 105(18): 6747-52.
von Kriegstein K, Eger E, Kleinschmidt A, Giraud AL. Modulation of neural responses to speech by directing attention to voices or verbal content. Brain ResCogn Brain Res 2003; 17(1): 48-55.
von Kriegstein K, Giraud AL. Distinct functional substrates along the right superior temporal sulcus for the processing of voices. Neuroimage 2004; 22(2): 948-55.
von Kriegstein K, Giraud AL. Implicit Multisensory Associations Influence Voice Recognition. PLoSBiol 2006; 4(10).
von Kriegstein K, Kleinschmidt A, Sterzer P, Giraud AL. Interaction of face and voice areas during speaker recognition. Journal of cognitive neuroscience 2005; 17(3): 367-76.
von Kriegstein K, Smith DR, Patterson RD, Kiebel SJ, Griffiths TD. How the human brain recognizes speech in the context of changing speakers. The Journal of neuroscience : the official journal of the Society for Neuroscience 2010; 30(2): 629-38.
von Kriegstein K, Warren JD, Ives DT, Patterson RD, Griffiths TD. Processing the acoustic effect of size in speech sounds. Neuroimage 2006; 32(1): 368-75.
Warren J, Scott S, Price C, Griffiths T. Human brain mechanisms for the early analysis of voices. Neuroimage 2006; 31(3): 1389-97.
Xu X, Biederman I, Shilowich BE, Herald SB, Amir O, Allen NE. Developmental phonagnosia: Neural correlates and a behavioral marker. Brain Lang 2015; 149: 106-17.

# Appendix

## List of Figures

### 2. Standard Model of Voice-Identity Processing



### 7 Study Findings Suggest Revised Version of Voice-Identity Processing Model

## Selbständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit ohne unzulässige Hilfe und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe und dass die aus fremden Quellen direkt oder indirekt übernommenen Gedanken in der Arbeit als solche kenntlich gemacht worden sind.

Claudia Roswandowitz

Leipzig, den 13. April 2017